On the relationship between model uncertainty and inferential/predictive uncertainty

BY DAVID DRAPER
Department of Applied Mathematics and Statistics, University of California, Santa Cruz, USA

SUMMARY

Inference and prediction are usually based on a model relating known and unknown quantities. The model itself typically contains features that are not known with certainty. In good applied work this is often dealt with by expanding a single "best" or standard model in ways that are sensitive to the model uncertainty, in effect embedding the single model in a larger family of models of which it is a special case. This represents a net gain in model uncertainty, and one might expect that uncertainty about the desired inference or prediction would increase as a result. But this need not be so; inferential/predictive uncertainty can either go up or down as model uncertainty increases. In this paper I discuss a common example in which inferential/predictive uncertainty decreases when model uncertainty goes up.

Key words: Calibration; Distributional uncertainty; Location-scale models; Minimal Fisher information; Modeling strategies; Serial correlation.

1. INTRODUCTION

We use models constantly in our inferential and predictive work. Most of the time, one or more features of such models are arrived at after a search (typically guided by the data) among possible modeling choices. Often we deal with the uncertainty in the modeling process implied by this search by ignoring it: we find the best model (in some sense) and carry out our inferences and predictions conditional on this model, as if it were "correct." Sometimes this produces well-calibrated answers (for instance, when there is little ambiguity about sensible modeling choices); sometimes it does not.

When we follow through to see if our inferential or predictive statements were right about as often as we asserted they would be, we find most frequently that lack of calibration is in the direction of insufficient conservatism—in other words, we had more uncertainty than we were willing to admit. One possible explanation for this is the underpropagation of modeling uncertainty just mentioned. One would ordinarily expect that acknowledging greater modeling uncertainty—for example, by making a standard modeling choice a special case of a larger family of models and enlisting the help of the data to search among this larger family for a plausible model, rather than just assuming the standard model is "correct"—would lead to an increase in inferential or predictive uncertainty. But this need not be so. In

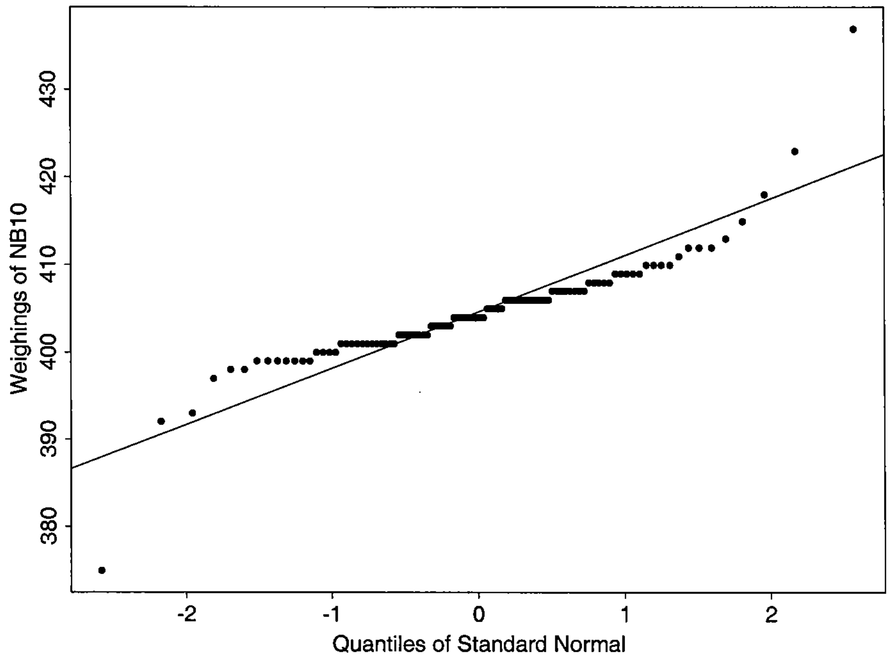

Fig. 1. Normal quantile-quantile plot of the NB10 data.

this paper I examine a common class of examples in which an increase in modeling uncertainty leads to a *decrease* in inferential uncertainty.

Figure 1 is a normal quantile-quantile plot of 100 weighings of a checkweight called NB10, made by workers at the U.S. National Bureau of Standards in 1962–63 under conditions as close to IID as possible (Freedman, Pisani, Purves & Adhikari, 1991; the units are micrograms below the nominal weight of 10g). How much does NB10 weigh? Figure 1 shows that it is plausible in answering this question to assume a symmetric location-scale model, but the form of the error distribution is less clear. The standard choice is Gaussian; with μ as the true weight and little or no prior information, the posterior distribution for μ based on an assumption of Gaussian errors is close to normal with mean 404.6 and standard deviation 0.65. But the solid line in Figure 1 is the target shape for the plot if the data were in fact Gaussian, and there is clear evidence for heavier tails. If one were to instead adopt (say) a t_k model for the errors and treat the degrees of freedom k as unknown—which corresponds to an increase in modeling uncertainty—it turns out that the posterior SD *drops* to 0.46, a 29% decrease. Is this a fluke or an example of a general phenomenon?

2. Increasing model uncertainty can decrease inferential uncertainty

Consider $Y_i, i = 1, \ldots, n$, IID (given (μ, σ)) from a symmetric location-scale family $Y_i = \mu + \sigma e_i$, where the e_i are assumed to have two finite moments and